

First mitochondrial genome from Yponomeutidae (Lepidoptera, Yponomeutoidea) and the phylogenetic analysis for Lepidoptera

Mingsheng Yang¹, Bingyi Hu¹, Lin Zhou¹, Xiaomeng Liu¹, Yuxia Shi¹,
Lu Song¹, Yunshan Wei², Jinfeng Cao³

1 College of Life Science and Agronomy, Zhoukou Normal University, Zhoukou, Henan, 466000, China

2 Chifeng Agricultural and Animal Husbandry Scientific Research Institute, Chifeng, Neimenggu, 024031, China

3 Cangzhou Academy of Agriculture and Forestry Sciences, Cangzhou, Hebei, 061001, China

Corresponding author: Mingsheng Yang (yms-888@163.com)

Academic editor: Thomas Simonsen | Received 2 April 2019 | Accepted 18 September 2019 | Published 9 October 2019

<http://zoobank.org/EEE86198-06CF-415A-A70D-F2944FBF5A52>

Citation: Yang M, Hu B, Zhou L, Liu X, Shi Y, Song L, Wei Y, Cao J (2019) First mitochondrial genome from Yponomeutidae (Lepidoptera, Yponomeutoidea) and the phylogenetic analysis for Lepidoptera. ZooKeys 879: 137–156. <https://doi.org/10.3897/zookeys.879.35101>

Abstract

The complete mitochondrial genome (mitogenome) of *Yponomeuta montanatus* is sequenced and compared with other published yponomeutoid mitogenomes. The mitogenome is circular, 15,349 bp long, and includes the typical metazoan mitochondrial genes (13 protein-coding genes, two ribosomal RNA genes, and 22 transfer RNA genes) and an A + T-rich region. All 13 protein-coding genes use a typical start codon ATN, the one exception being *cox1*, which uses CGA across yponomeutoid mitogenomes. Comparative analyses further show that the secondary structures of tRNAs are conserved, including loss of the Dihydropyrimidine (DHU) arm in *trnS1* (AGN), but remarkable nucleotide variation has occurred mainly in the DHU arms and pseudouridine (T ψ C) loops. A + T-rich regions exhibit substantial length variation among yponomeutoid mitogenomes, and conserved sequence blocks are recognized but some of them are not present in all species. Multiple phylogenetic analyses confirm the position of *Y. montanatus* in Yponomeutoidea. However, the superfamily-level relationships in the Macroheterocera clade in Lepidoptera recovered herein show considerable difference with that recovered in previous mitogenomic studies, raising the necessity of extensive phylogenetic investigation when more mitogenomes become available for this clade.

Keywords

Mitogenome evolution, next-generation sequencing, protein-coding genes

Introduction

The mitochondrial genome (mitogenome) is a circular and double-stranded molecule that usually encodes 37 genes (13 protein-coding genes (PCGs), two ribosomal RNA genes (rRNAs), 22 transfer RNA genes (tRNAs)), and an A + T-rich region (Boore 1999). Characterized by cellular abundance, absence of introns, rapid evolution, and a lack of extensive recombination, mitogenome sequences can be easily amplified and has been extensively employed in evolutionary studies in past decades (Cameron et al. 2014; Curole et al. 2014). Additionally, the mitogenome often exhibits some characters such as gene rearrangement that have been widely used to infer genome evolution and phylogeny for multiple groups. For instance, comparative analyses of lepidopteran mitogenomes showed the gene order *trnM-trnI-trnQ* is present in more derived ditrysian lineage and its close relatives Tischerioidea and Palaephatoidea, in contrast to the ancestral *trnI-trnQ-trnM* in other lepidopterans such as *Thitarodes renzhiensis* (Yang, 1991) and *T. yunnanensis* (Yang, 1992) (Cao et al. 2012; Timmermans et al. 2014).

Lepidoptera is the second largest insect order after Coleoptera, with more than 157,000 extant species in 43 superfamilies (van Nieukerken et al. 2011; Mitter et al. 2017). To date, mitogenomes of more than 400 lepidopteran species or subspecies have been sequenced (<https://www.ncbi.nlm.nih.gov>; last visited on March 2019). Relative to other species-rich orders, however, the current number of sequenced mitogenomes is still limited. Moreover, deep-level lepidopteran phylogeny is still poorly understood despite previous investigations based on various data including mitogenome sequences (Mitter et al. 2017). The superfamily Yponomeutoidea, with approximately 1,800 described species, represents one of the earliest diverging lineages of ditrysian Lepidoptera and includes many notable pest species (van Nieukerken et al. 2011; Sohn et al. 2013). In Yponomeutoidea, 11 families were recognized based on a multiple-gene dataset (Sohn et al. 2013), but phylogenetic relationships among yponomeutoid families still need further investigation (Mitter et al. 2017). To date, mitogenomes of only three yponomeutoid species representing three families have been published. According to the classification system proposed by Sohn et al. (2013), they are *Prays oleae* (Bernard, 1788) (Praydidae) (van Asch et al. 2016), *Plutella xylostella* (Linnaeus, 1758) (Plutellidae) (Wei et al. 2013; Dai et al. 2016), and *Leucoptera malifoliella* (Costa, 1836) (Lyonetiidae) (Wu et al. 2012). Thus, the number of reported yponomeutoid mitogenomes is quite limited. Moreover, a comparative analysis among the published mitogenomes has never been conducted.

Mitogenomic data of major Yponomeutoidea lineages would play an important role for better understanding the evolution of the superfamily or even Lepidoptera as a whole. In the present study, we sequenced the complete mitogenome of *Yponomeuta montanatus* Moriuti, 1977, the first mitogenome from the family Yponomeutidae. Moreover, detailed comparative analyses were conducted based on this and all other published yponomeutoid mitogenomes. In addition, extensive phylogenetic analyses using three different datasets and three different tree-constructed methods were performed to test phylogenetic implications of the *Y. montanatus* mitogenome in Lepidoptera phylogeny. This study contributes to further understanding the mitogenome evolution and phylogeny of the Yponomeutoidea and Lepidoptera.

Materials and methods

Sample collection, identification and DNA extraction

Adult *Y. montanatus* specimens were sampled by light trap at Mountain Jigongshan, Henan, China in May 2018. Fresh specimens were stored in 95–100% ethanol in the field and then maintained at -80°C until used for DNA extraction. Dry specimens were identified based on the morphological description and illustrations provided by Byun and Bae (2013). In addition, molecular identification was performed by blasting the standard *cox1* barcode sequence in GenBank. Thorax muscle tissues were used to extract genomic DNA with the DNeasy tissue kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Voucher specimens are deposited in the Biology Laboratory of Zhoukou Normal University, China.

Mitogenome sequencing and assembly

Next-generation sequencing methods were used to obtain the complete mitogenome sequence of *Y. montanatus*. Briefly, total genomic DNA was firstly quantified and fragmented to an average size of 400 bases using Covaris M220 system with the Whole Genome Shotgun method (Covaris, Woburn, MA, USA). Then, a library was constructed using the TruSeq DNA PCR-Free Sample Preparation Kit (Illumina, USA). Lastly, Illumina HiSeq 2500 was used for sequencing with the strategy of 251 paired-ends.

A total of 3,707,876 raw paired reads were retrieved for *Y. montanatus*. FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) was used for quality control (avg. Q20 > 95.1%, avg. Q30 > 88.65%). After processing with AdapterRemoval v. 2 (Mikkelsen et al., 2016) and SOAPdenovo v. 2.01 (Luo et al. 2012), the raw paired reads were filtrated into a total of 2,582,644 high-quality reads. Then, the A5-miseq v20150522 (Coil et al. 2015) and SPAdes v. 3.9.0 (Bankevich et al. 2012) were used in de novo assembly, generating contig and scaffold sequences. Lastly, the mitochondrial sequences were identified using blastn method, and the mummer v. 3.1 (Kurtz et al. 2004) was used to establish the position relationships among the contig sequences and to fill in the possible gaps.

Mitogenome annotation and comparative analysis

The MITOS webserver was employed to annotate the complete mitogenome sequence with the invertebrate genetic code (Bernt et al. 2013). The tRNAScan-SE server v. 1.21 (Lowe and Eddy 1997) was used to re-identify the 22 tRNAs as well as to reconfirm their secondary structures. Gene boundaries were re-identified by aligning the new mitogenome with previously reported yponomeutoid mitogenomes using MEGA v. 6.06 (Tamura et al. 2013). To ensure the correct reading frame, nucleotide sequences of the 13 PCGs were translated with both the programs Primer Premier v. 5.00 (Pre-

mier Biosoft International, Palo Alto, CA) and MEGA v. 6.06 (Tamura et al. 2013) with invertebrate mitochondrial genetic code. Tandem repeat elements in the A + T-rich region were identified using the Tandem Repeats Finder program (<http://tandem.bu.edu/trf/trf.html>) (Benson 1999). All other published yponomeutoid mitogenomes, along with the one sequenced in this study were compiled for comparative analysis. Base composition and the relative synonymous codon usage (RSCU) were calculated using MEGA v. 6.06 (Tamura et al. 2013). Strand asymmetry was calculated according to the formulas: $AT\text{-skew} = [A - T]/[A + T]$ and $GC\text{-skew} = [G - C]/[G + C]$ (Perna and Kocher 1995). The nucleotide diversity and the ratio of non-synonymous substitution (K_a) to synonymous substitution (K_s) were calculated with DNASP v. 5.0 (Librado and Rozas 2009).

Phylogenetic analyses

To investigate phylogenetic implications of the *Y. montanatus* mitogenome in Lepidoptera phylogeny, a total of 33 mitogenomes representing 15 lepidopteran superfamilies with mitogenome available (Suppl. material 1, Table S1) were sampled for phylogenetic analyses. Two additional trichopteran mitogenomes were selected as outgroups. Sequence alignment was conducted on the TranslatorX online platform (Abascal et al. 2010) for 13 PCGs. The two rRNAs and 22 tRNAs were aligned with the Q-INS-i algorithm implemented in the MAFFT online platform (Katoh et al. 2017). MEGA v. 6.06 (Tamura et al. 2013) was used to check all alignments. Then, MEGA v. 6.06 (Tamura et al. 2013) was also used to generate three different datasets: PCG123 (all codon positions of 13 PCGs), PCG123R (PCG123 dataset plus two rRNAs and 22 tRNAs), and PCGAA (amino acid sequences translated from 13 PCGs). Nucleotide sequence substitution model was selected using PartitionFinder v. 1.1.1 (Lanfear et al. 2012), with the Bayesian Information Criterion (BIC) algorithm under a greedy search. The best partition scheme and corresponding models are shown in Suppl. material 1, Tables S2 and S3.

Maximum likelihood (ML) analyses were conducted using two methods. The raxmlGUI version 1.539 interface (Silvestro and Michalak 2012) of RAxML version 7.2.6 (Stamatakis 2006) was used under the GTRGAMMAI model for PCG123 and PCG123R datasets, and the model MtArt + I + G for PCGAA dataset. Node reliability was assessed using the ML + rapid bootstrap algorithm with 100 replicates. IQ-TREE 1.6.7.1 (Nguyen et al. 2015) was used with the models determined by PartitionFinder for PCG123 and PCG123R datasets, and the model MtArt + I + G for PCGAA dataset. Node support was assessed using 1,000 ultrafast bootstrap replicates.

Bayesian inference (BI) analysis was performed using MrBayes v. 3.1.2 (Ronquist and Huelsenbeck 2003). For the PCG123 and PCG123R datasets, the model determined by PartitionFinder was used, and for PCGAA dataset with the model MtRev + I + G. Two independent Markov chain Monte Carlo (MCMC) runs were performed for 1,000,000–3,000,000 generations sampling per 100 generations. The convergence

between the two runs was established by the Tracer v. 1.6 (Effective sample sizes >200) (Rambaut et al. 2014). After the first 25% of the yielded trees were discarded as burn-in, a 50% majority-rule consensus tree with the posterior probability was generated from the remaining trees.

Results and discussion

General characteristics of the *Y. montanatus* mitogenome

The complete mitogenome of *Y. montanatus* (GenBank accession number: [MK256747](#)) is circular, double-stranded, and 15,349 bp long (Fig. 1, Table 1). This length is shortest amongst published yponomeutoid mitogenomes. The typical 37 mitochondrial genes (13 PCGs, 22 tRNAs, and two rRNAs) and an A + T-rich region are included. Among them, 23 (nine PCGs and 14 tRNAs) are encoded on majority strand (J-strand), and the remaining 14 are located on minority strand (N-strand). As in most ditrysian members of Lepidoptera, the *trnM-trnI-trnQ* can be recognized in Yponomeutoidea, in contrast to the *trnI-trnQ-trnM* in most non-ditrysian lineage such as the Hepialoidea (Cao et al. 2012), and in the ancestral arthropod mitogenome (*Drosophila yakuba*) (Clary and Wolstenholme 1985).

As in other insect mitogenomes (Boore 1999), high A + T content is recognized across Yponomeutoidea mitogenomes, which ranges from 81% in *P. lutella* ([KM023645](#)) to 82.5% in *L. malifoliella* (Table 2). In addition to the A + T content, AT-skew and GC-skew are also routinely used to characterize base composition of mitogenomes (Perna and Kocher 1995; Wei et al. 2010). The negligible AT-skew (0.0037) and moderate GC-skew (−0.164) (Suppl. material 1, Table S4) in *Y. montanatus* mitogenome are similar to other Lepidoptera and most insect species (Cameron and Whiting 2008).

Protein-coding genes

The total length of the 13 PCGs in *Y. montanatus* mitogenome is 11,183 bp, approximately accounting for 72.9% of the whole mitogenome (Table 2). Identical to other yponomeutoid mitogenomes, nine of the 13 PCGs are encoded on J-strand, and the other four are located on N-strand. In yponomeutoid mitogenomes, the A + T content of the 13 PCGs varies from 79.1% in *P. oleae* to 80.7% in *L. malifoliella*. The codon positions show unequal A + T content (Suppl. material 1, Table S5). The third codon positions have the highest A + T content (93.4% on average), followed by first codon positions (74.9% on average) and second codon positions (70.7% on average). To characterize codon frequencies across yponomeutoid mitogenomes, relative synonymous codon usages (RSCU) were calculated and drawn for all five yponomeutoid mitogenomes. As shown in Figure 2 and Suppl. material 1, Table S6, the codon usage

Table 1. Summary of the *Yponomeuta montanatus* mitogenome.

Feature	Strand	Location	Size (bp)	Start codon	Stop codon	Anticodon	Intergenic nucleotides
<i>trnM</i>	J	1–67	67			CAT	0
<i>trnI</i>	J	68–136	69			GAT	–3
<i>trnQ</i>	N	134–202	69			TTG	49
<i>nad2</i>	J	252–1265	1014	ATT	TAA		–2
<i>trnW</i>	J	1264–1331	68			TCA	–8
<i>trnC</i>	N	1324–1385	62			GCA	11
<i>trnY</i>	N	1397–1460	64			GTA	2
<i>cox1</i>	J	1463–2998	1536.9	CGA	TAA		–5
<i>trnL2</i> (UUR)	J	2994–3059	66			TAA	0
<i>cox2</i>	J	3060–3744	685	ATG	T		–3
<i>trnK</i>	J	3742–3812	71			CTT	–1
<i>trnD</i>	J	3812–3876	65			GTC	0
<i>atp8</i>	J	3877–4035	159	ATT	TAA		–7
<i>atp6</i>	J	4029–4706	678	ATG	TAA		–1
<i>cox3</i>	J	4706–5497	792	ATG	TAA		2
<i>trnG</i>	J	5500–5565	66			TCC	0
<i>nad3</i>	J	5566–5919	354	ATT	TAA		2
<i>trnA</i>	J	5922–5984	63			TGC	–1
<i>trnR</i>	J	5984–6050	67			TCG	8
<i>trnN</i>	J	6059–6123	65			GTT	–1
<i>trnS1</i> (AGN)	J	6123–6188	66			GCT	0
<i>trnE</i>	J	6189–6250	62			TTC	–1
<i>trnF</i>	N	6250–6315	66			GAA	22
<i>nad5</i>	N	6338–8050	1713	ATT	TAA		12
<i>trnH</i>	N	8063–8129	67			GTG	0
<i>nad4</i>	N	8130–9468	1339	ATG	T		0
<i>nad4I</i>	N	9469–9756	288	ATG	TAA		7
<i>trnT</i>	J	9764–9828	65			TGT	0
<i>trnP</i>	N	9829–9894	66			TGG	2
<i>nad6</i>	J	9897–10430	534	ATT	TAA		9
<i>cob</i>	J	10440–11591	1152	ATG	TAA		–2
<i>trnS2</i> (UCN)	J	11590–11657	68			TGA	35
<i>nad1</i>	N	11693–12631	939	ATG	TAA		1
<i>trnL1</i> (CUN)	N	12633–12699	67			TAG	0
<i>rrnL</i>	N	12700–14073	1374				0
<i>trnV</i>	N	14072–14135	64			TAC	–1
<i>rrnS</i>	N	14135–14903	769				0
A + T-rich region		14904–15349	446				

Note: the “J” indicates the majority strand and the “N” indicates the minority strand in the strand column.

pattern is generally similar among them such as the most frequently used codons (i.e., UUA, AUU, UUU, AUA, and AAU). In the *Y. montanatus* mitogenome, a number of 3,727 amino acids are translated, of which 1,787 (47.9%) are encoded by the five frequently used codons above. However, the codons absent in yponomeutoid mitogenomes are different, but most of them are rich in C/G nucleotides. In general, the high A/T content in frequently used codons effectively contributes to the high A + T composition in PCGs and the whole mitogenome.

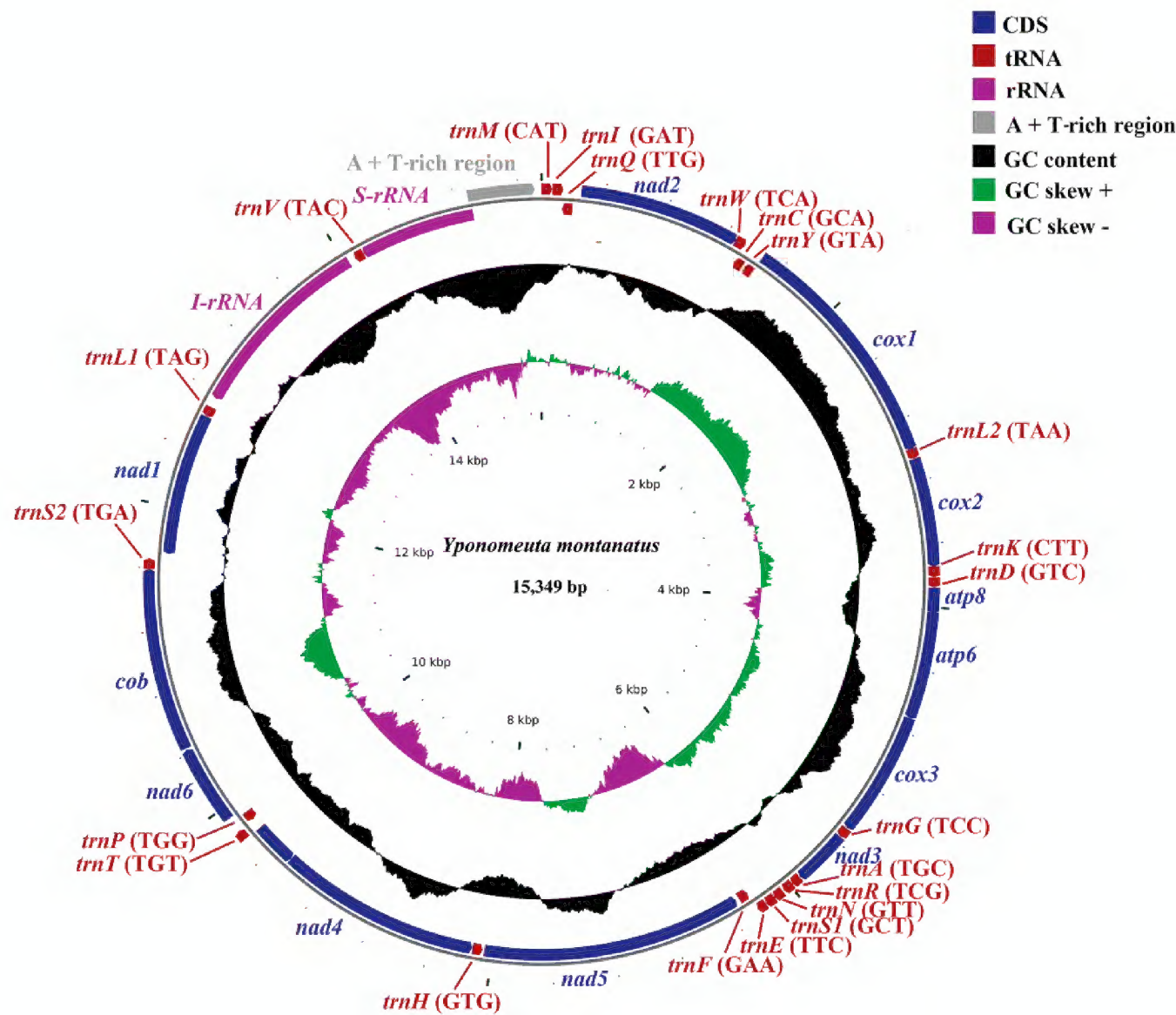


Figure 1. Mitochondrial genome map of the *Yponomeuta montanatus*.

Table 2. Base composition of the five sequenced yponomeutoid mitogenomes.

Taxon	Size (bp)	A + T (%)	PCGs No. of codon A + T (%)	<i>rrnS</i> RNA Size (bp) A + T (%)	<i>rrnL</i> RNA Size (bp) A + T (%)	tRNAs Size (bp) A + T (%)	A + T-rich region Size (bp) A + T (%)	GenBank accession no.
Yponomeutidae								
<i>Yponomeuta montanatus</i>	15,349	81.08	3,727 79.6	769 85.7	1,374 85.1	1,453 80.8	446 96.2	MK256747
Praydidae								
<i>Prays oleae</i>	16,499	81.8	3,720 79.1	773 85	1,372 85	1,486 81.3	1,483 96.3	KM874804
Plutellidae								
<i>Plutella xylostella</i>	16,014	81	3,731 79.4	783 86.1	1,382 85.1	1,465 81.2	888 93.1	KM023645
<i>Plutella xylostella</i>	16,179	81.4	3,729 79.4	783 86.1	1,415 84.9	1,468 81.3	1,081 n.a.	JF911819
Lyonetiidae								
<i>Leucoptera malifoliella</i>	15,646	82.5	3,719 80.7	770 87.1	1,351 85.5	1,488 83.7	733 95.3	JN790955

Note: n.a. indicates not available.

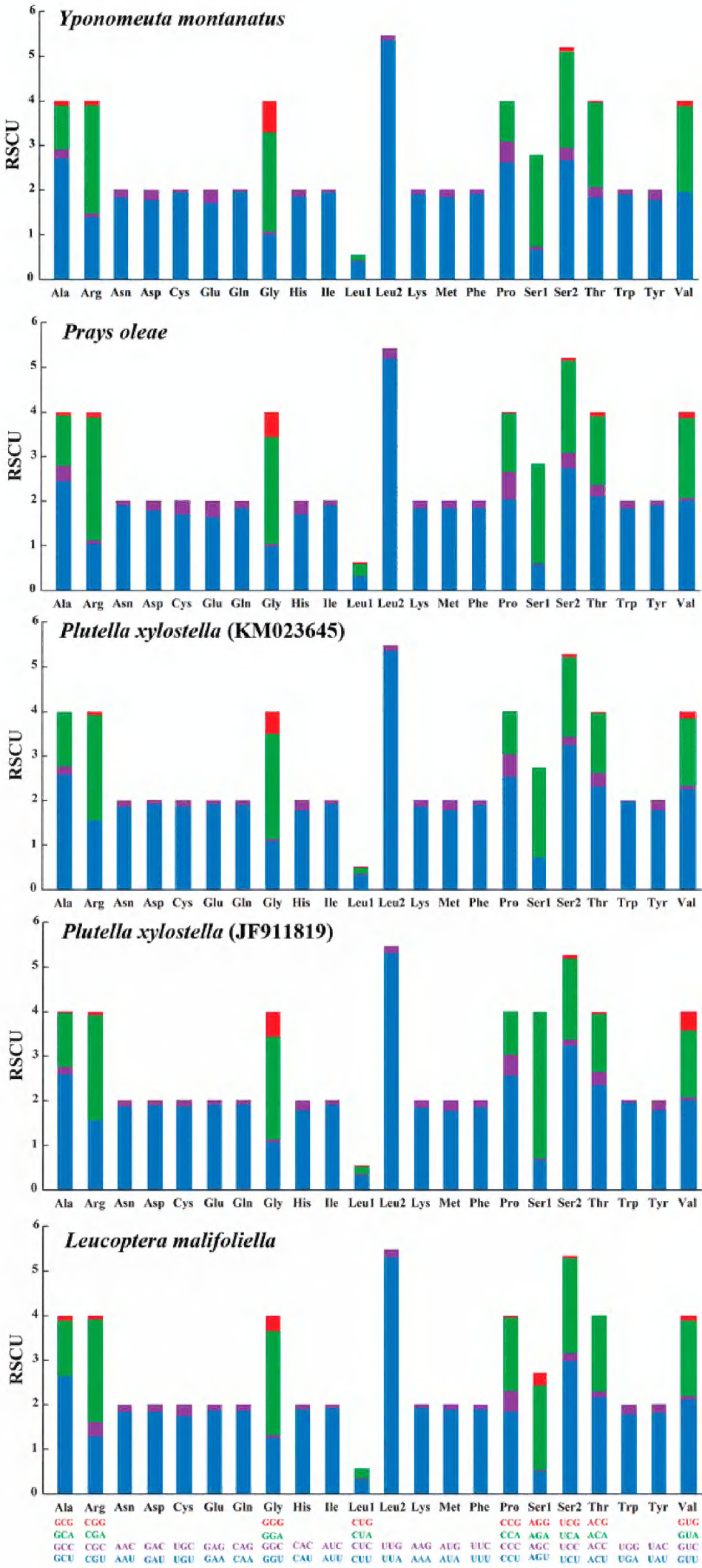


Figure 2. Relative synonymous codon usages (RSCU) in PCGs of *Yponomeuta montanatus* and other published yponomeutoid mitogenomes. Codon families are indicated below the X-axis.

Most PCGs in yponomeutoid mitogenomes use the conventional ATN start codon (Table 1, Suppl. material 1, Table S7). The unconventional CGA was consistently found in only the *cox1* gene, a common feature for lepidopteran mitogenomes (Wu et al. 2016). TAA is employed as stop codon in most PCGs, but two other kinds of stop codons were recognized. One is the TAG for *nad4l* and *nad6* genes in *L. malifoliella*; the other is the incomplete termination codon T which is commonly used in yponomeutoid mitogenomes. Actually, the incomplete termination codon can be also commonly recognized across arthropod mitogenomes, which may be related to the post transcriptional modification during the mRNA maturation process (Ojala et al. 1981).

To investigate evolutionary patterns of all PCGs, nucleotide diversity and the ratio of Ka to Ks were calculated for each PCG. As shown in Figure 3 and Suppl. material 1, Table S8, *nad6* and *cox1* genes show the highest and lowest nucleotide diversity respectively. The Ka/Ks values are the highest in *atp8* genes and the lowest is for *cox1* gene. Notably, the Ka/Ks values for all PCGs are lower than one, indicating that they are evolving under the purifying selection and are suitable for investigating phylogenetic relationships within Yponomeutoidea.

Transfer and ribosomal RNA genes

The *Y. montanatus* mitogenome contains 22 tRNAs with the length ranging from 62 bp (*trnC*, *trnE*) to 71 bp (*trnK*) (Fig. 4, Table 1). Among them, eight tRNAs are encoded by N-strand and the remaining 14 by J-strand. The total length of tRNAs is 1,453 bp, which is shortest among yponomeutoid mitogenomes, which otherwise range from 1,465 bp in *P. xylostella* (KM023645) to 1,488 in *L. malifoliella* (Table 2). As shown in Fig. 4, all tRNAs exhibit typical clover-leaf secondary structure but *trnS1* (AGN) lacks the DHU arm, a feature generally present in all Lepidoptera insects as well as in other metazoan mitogenomes (Garey and Wolstenholme 1989; Lavrov et al. 2000). In tRNAs of the *Y. montanatus* mitogenome, we recognized 22 unmatched base pairs, of which 18 are non-canonical G-U pairs, and the remaining four are mis-

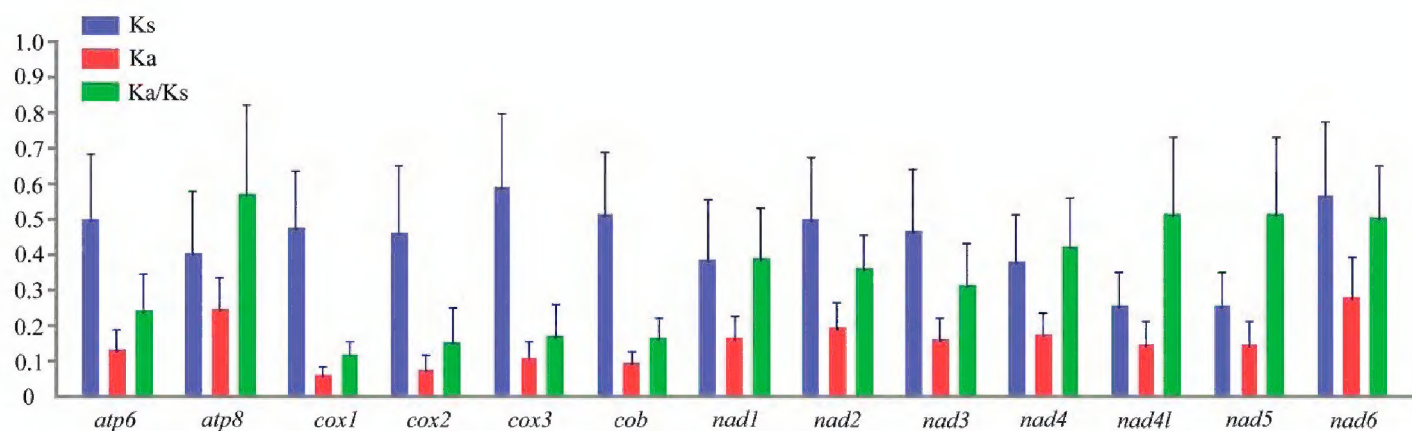


Figure 3. Evolutionary rate of each PCG among yponomeutoid mitogenomes. Ka, non-synonymous substitution; Ks, synonymous substitution.

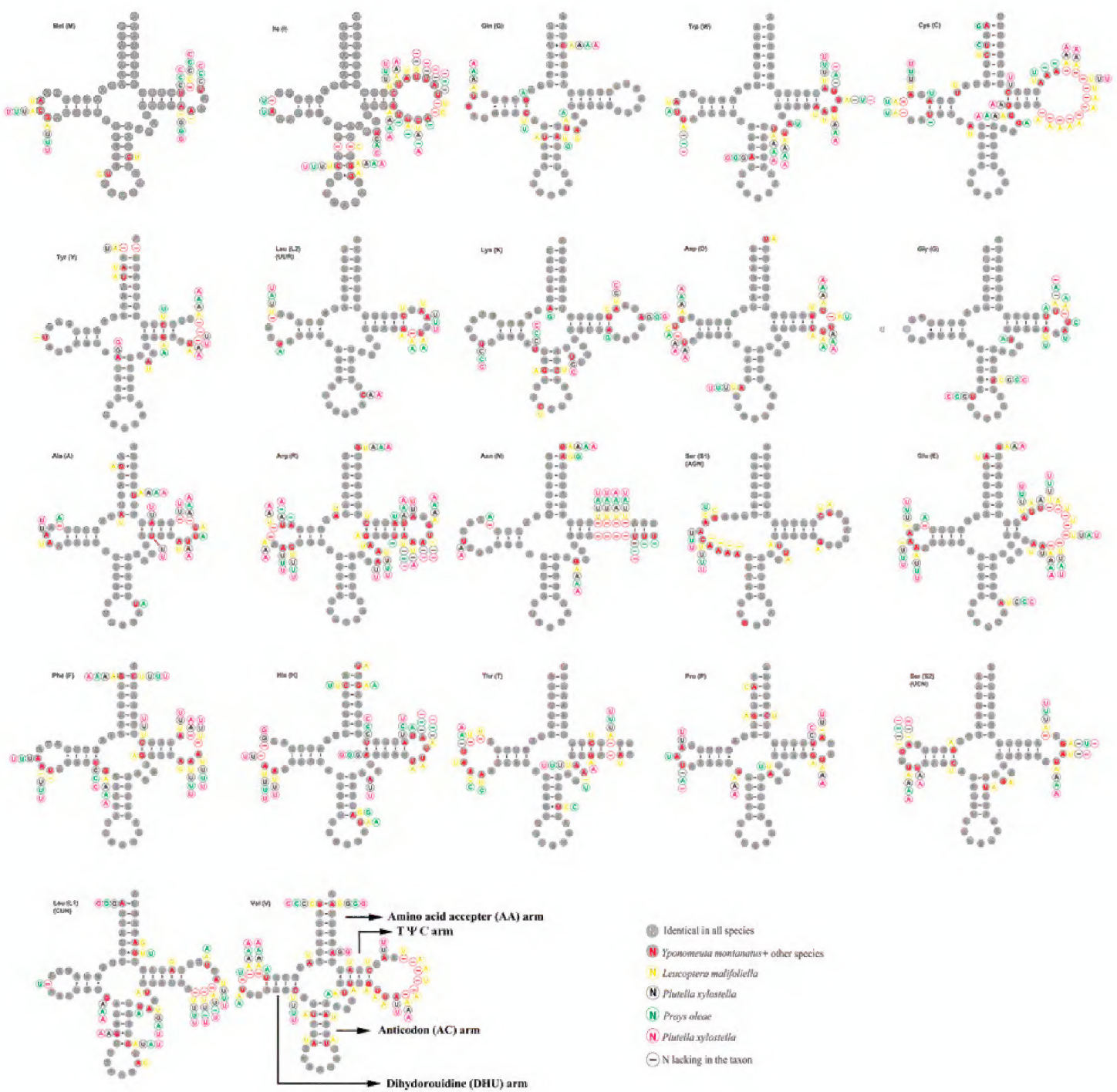


Figure 4. Putative secondary structures of tRNAs from *Yponomeuta montanatus* mitogenome. The tRNAs are labeled with the abbreviations of their corresponding amino acids. The tRNA arms are illustrated as for *trnV*. Dashes indicate the Watson-Crick base pairs; dots indicate the wobble GU pairs; and the other non-canonical pairs are not marked. The nucleotides marked indicate the variable sites among published yponomeutoid mitogenomes.

matched base pairs U-U. The overrepresented non-canonical G-U pairs in tRNAs is commonly present in insect mitogenomes (Salvato et al. 2008; Chen et al. 2016; Chen and Du 2017).

Comparative tRNA analyses among yponomeutoid mitogenomes found that each tRNA structure is highly conserved, including the loss of the DHU arm in *trnS1* (AGN). However, substantial nucleotide variation exists, most of which occurred in the DHU arm and TΨC loops (Fig. 4). Interestingly, in *L. malifoliella*, the anticodons for both *trnK* and *trnS1* (AGN) were rarely mutated relative to other yponomeutoid

species. For *trnS1* (AGN), the TCT is used instead of routinely used codon GCT. This phenomenon has been recognized in previous reports such as two species of Heptaloidea and three species of Noctuoidea (Li et al. 2018). In *trnK*, the mostly used anticodon CTT was changed to TTT (Wu et al. 2012), which is, to our knowledge, extremely rare in Lepidoptera and insects in general.

Similar to other yponomeutoid mitogenomes, two rRNA genes, *rrnS* and *rrnL*, were recognized in the *Y. montanatus* mitogenome (Fig. 1, Table 1). The *rrnS* is 769 bp long, which is located between *trnV* and A + T-rich region; the *rrnL* is 1,374 bp long, being present between *trnV* and *trnL1*. The lengths of *rrnS* and *rrnL* are comparable to those of other reported yponomeutoid mitogenomes, which are from 770 bp in *L. malifoliella* to 783 bp in *P. xylostella*, and from 1,351 bp in *L. malifoliella* to 1,415 bp in *P. xylostella* (JF911819), respectively.

Gene overlapping and intergenic regions

In the *Y. montanatus* mitogenome, 36 gene overlapping sites were recognized across 13 gene junctions from one to eight bp in length (Table 1). Comparative mitogenome analyses showed that gene overlapping region only between *atp8* and *atp6* is consistently present across reported yponomeutoid mitogenomes. This 7-bp motif of “ATGA-TAA” (Fig. 5A) is actually a common feature for Lepidoptera and other insects, such as *Taeniopteryx ugola* Ricker & Ross, 1968 (Plecoptera) (Chen and Du 2017).

In addition to the A + T-rich region, a total of 162 intergenic nucleotides across 13 gene junctions from one to 49 bp were identified in *Y. montanatus* mitogenome (Table 1). Among the 13 intergenic regions or site, three are conserved among the reported yponomeutoid mitogenomes, and they are located between the *nad6* and *cob* genes (Fig. 5B), the *trnQ* and *nad2* genes (Fig. 5C), and the *trnS2* and *nad1* genes (Fig. 5D). The one between the *nad6* and *cob* genes ranges from four to 41 bp in length across yponomeutoid mitogenomes. A remarkable feature for this region is that both mitogenomes of *P. xylostella* contain microsatellite (TA)_n sequence but with different repeat numbers. *P. xylostella* is an important agricultural pest, and this microsatellite (TA)_n sequence may be used as a candidate marker to test the population structure for pest management. The one between the *trnQ* and *nad2* genes exhibits substantial sequence variation (except two sequences for *P. xylostella*) among reported yponomeutoid mitogenomes. This intergenic region is also widely present in other lepidopterans such as those in the Lasiocampidae (Kim et al. 2017) and Noctuidae (Chai and Du 2012), and may even be regarded as an autapomorphy of Lepidoptera (Cao et al. 2014). Also, the intergenic region between *trnS2* and *nad1* is commonly present in insect mitogenomes (Cameron and Whiting 2008). Although the length varies among yponomeutoid mitogenomes, a conserved motif “ATACTAA” could be identified, which has been reported related to mitochondrion transcription (Taanman 1999).

“ATAG” and subsequent poly-T structure, the motif “ATTTA” and followed microsatellite (TA)_n element and an “A”-rich 3’ end upstream of the *trnM* gene. However, in *L. malifoliella* and *P. xylostella*, we did not recognize the poly-T structure and microsatellite (TA)_n element, respectively. Also, insect A + T-rich region is generally characterized by the presence of multiple tandem repeat elements (Vila and Björklund 2004). In yponomeutoid mitogenomes, this character can be recognized in *P. oleae* (van Asch et al. 2016) and *L. malifoliella* (Wu et al. 2012). However, in mitogenomes of *Y. montanatus* sequenced herein and *P. xylostella* (Dai et al. 2016), no tandem repeat elements were identified.

Phylogenetic analyses

To investigate phylogenetic implications of the *Y. montanatus* mitogenome in Yponomeutoidea and Lepidoptera, we constructed the superfamily-level relationships within Lepidoptera using three inference methods and three different datasets.

As shown in Figures 6–8. and S1–S3, relationships among the four yponomeutoid families involved herein were consistently recovered as Lyonetiidae + (Praydidae + (Yponomeutidae + Plutellidae)), which is consistent with that of Sohn et al. (2013) based on multiple-gene data. *Y. montanatus* is nested within Yponomeutoidea, confirming its phylogenetic position using mitogenomic data. In previous studies, the Yponomeutoidea is recovered either sister to Gracillarioidea (Heikkilä et al. 2015) or paraphyletic with respect to Gracillarioidea (Regier et al. 2013; Sohn et al. 2013). Most mitogenome-based phylogenetic studies of Lepidoptera scarcely sampled representatives of Gracillarioidea. As an exception, Timmermans et al. (2014) revealed that Gracillarioidea are nested in the Yponomeutoidea. The same results are recovered in this study. The only representative of Gracillarioidea is consistently sister to *L. malifoliella* in Yponomeutoidea, rendering the Yponomeutoidea paraphyletic.

Regarding the phylogenetic pattern of other superfamilies, mostly identical results were obtained by different analyses, which are also similar to other mitogenome-based studies (Clary and Wolstenholme 1985; Timmermans et al. 2014; Bao et al. 2018). We noticed that the minor topology difference across our analyses mainly occurred in the position of the Papilionoidea, Gelechioidea and Cossoidea. These results are similar to other mitogenome-based studies (Clary and Wolstenholme 1985; Bao et al. 2018) as well as multiple-gene-based study (Heikkilä et al. 2015) where the positions of these superfamilies are unstable or weakly supported, respectively. Within Macroheterocera, Noctuoidea + (Geometroidea + Bombycoidea) were recovered by studies based on various data, such as mitogenome sequences (Kim et al. 2011; Yang et al. 2015), multi-gene sequences (Regier et al. 2013) and 741 genes from transcriptome sequences (Bazinet et al. 2013). Interestingly, our analyses consistently recovered Bombycoidea + (Geometroidea + Noctuoidea), which was identical to that of Kawahara and Breinholt (Kawahara and Breinholt 2014). This result suggests the necessity of extensive phylogenetic investigation when more mitogenomes become available for this clade.

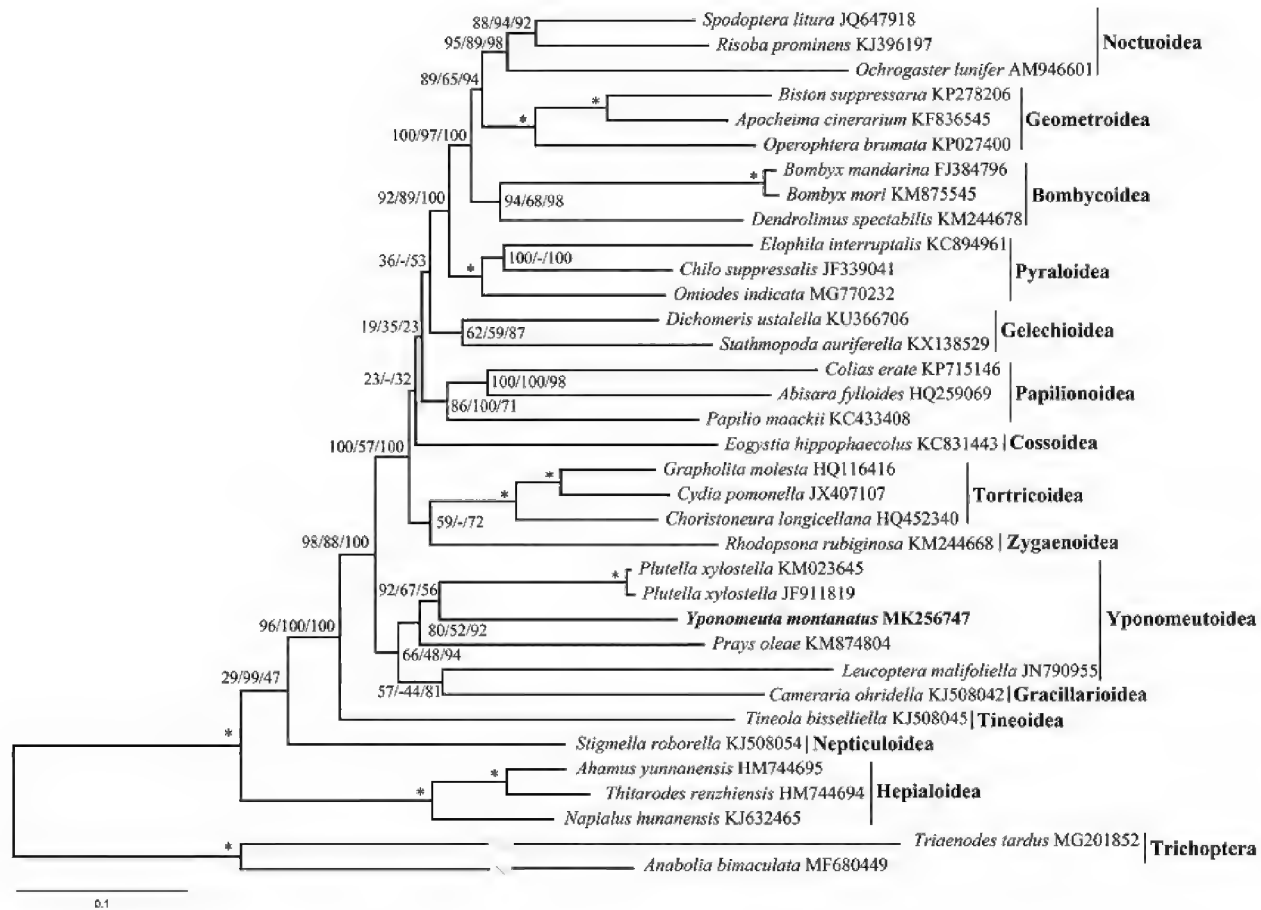


Figure 6. ML tree inferred from RAxML method based on PCG123R dataset. Numbers separated by slash (/) on node represent bootstrap replicates based on PCG123, PCGAA and PCG123R datasets, respectively. The dash (-) represents unrecovered node in ML tree based on the PCG123 or PCGAA dataset.

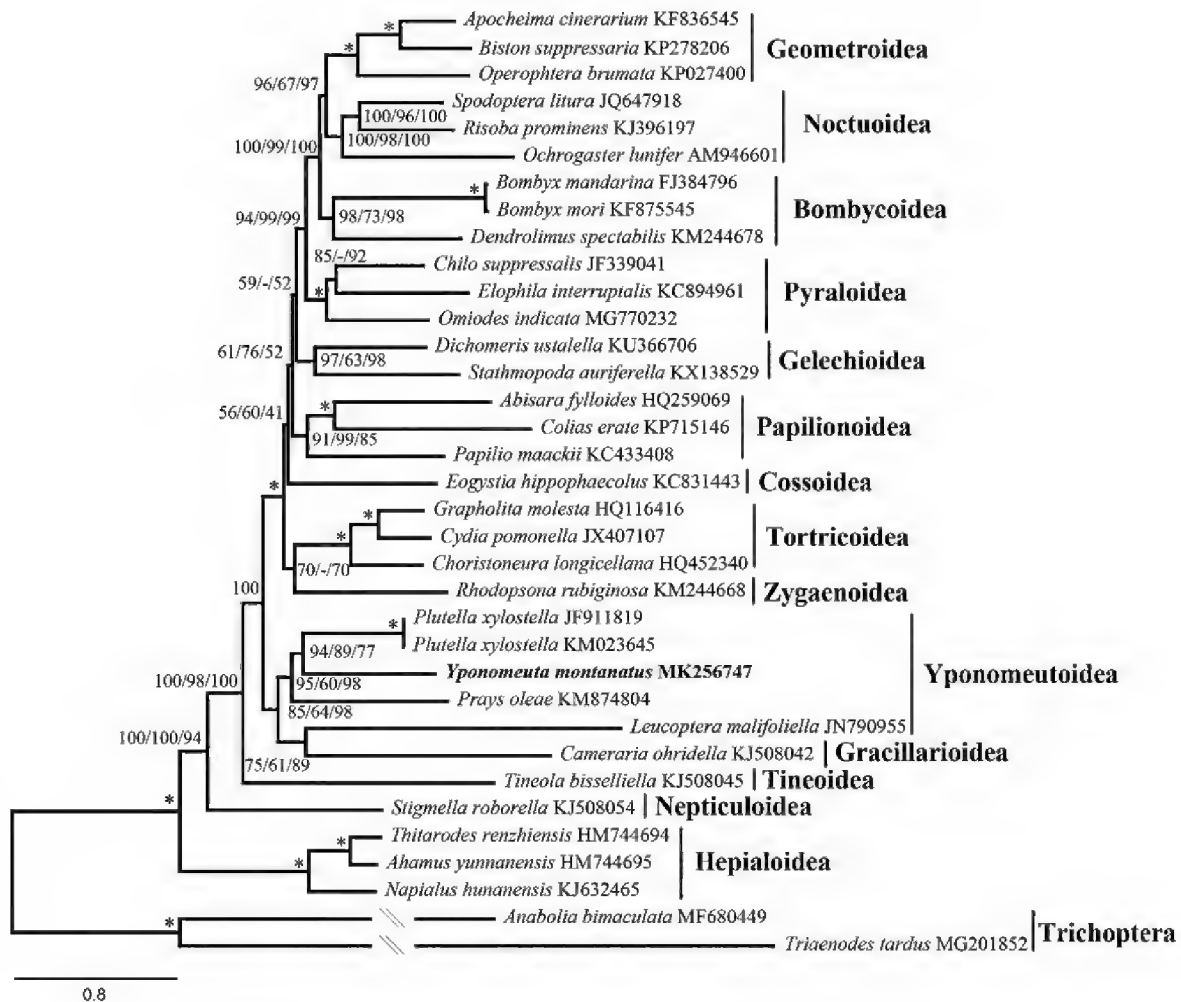


Figure 7. ML tree inferred from IQ-TREE method based on PCG123R dataset. Numbers separated by slash (/) on node represent bootstrap replicates based on PCG123, PCGAA and PCG123R datasets, respectively. The dash (-) represents unrecovered node in ML tree based on the PCG123 or PCGAA dataset.

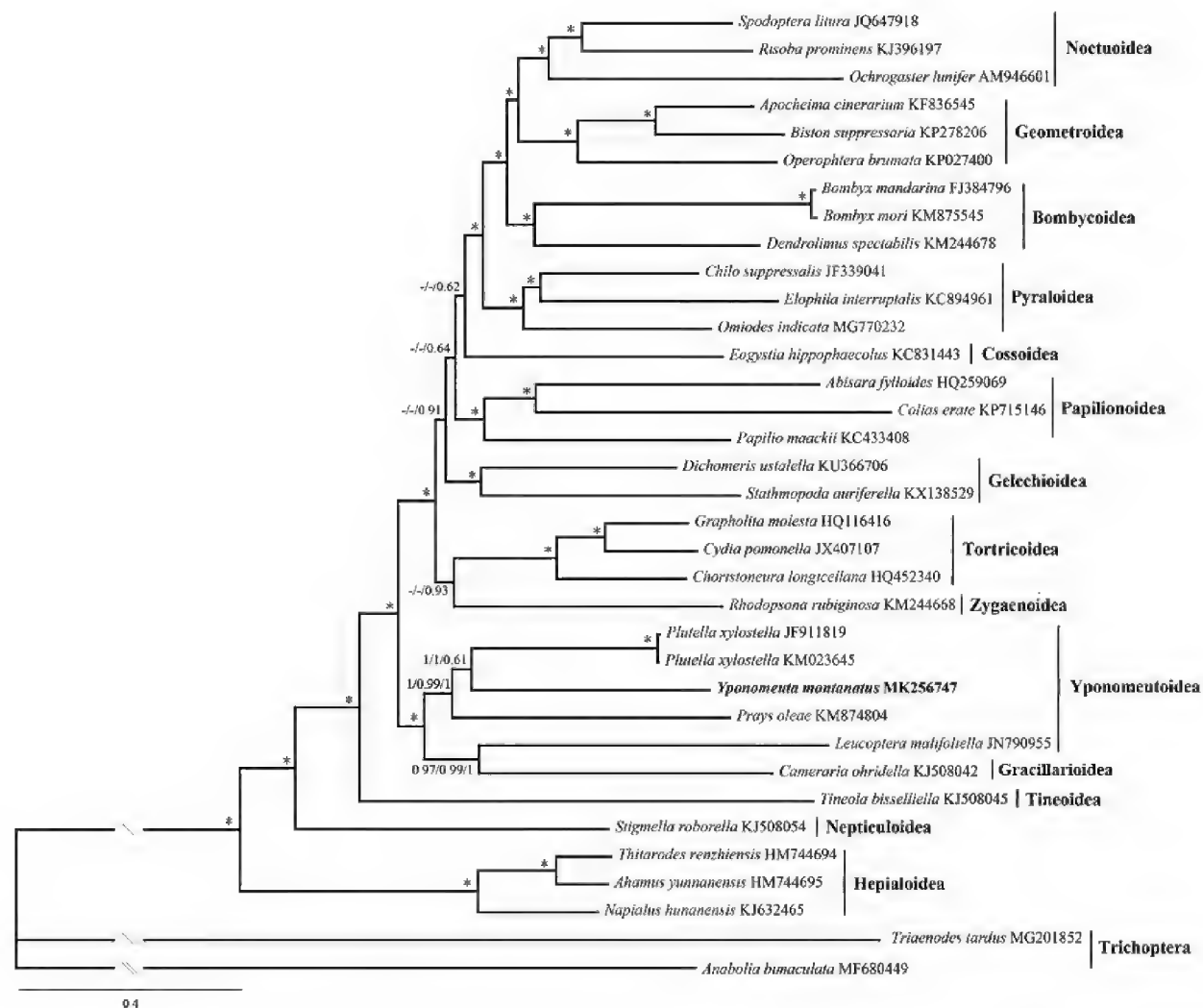


Figure 8. BI tree inferred from MrBayes method based on PCG123R dataset. Numbers separated by slash (/) on node represent posterior probabilities based on PCG123, PCGAA and PCG123R datasets, respectively. The dash (-) represents unrecovered node in BI tree based on the PCG123 or PCGAA dataset.

Acknowledgements

We sincerely appreciate the anonymous reviewers for comments on this manuscript. This work was supported by the National Natural Science Foundation of China (31702046), Scientific and Technological Innovation Talent Project of Henan Province (19HASTIT015) and Science and Technology Program of Henan Province (182106000047).

References

- Abascal F, Zardoya R, Telford MJ (2010) TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research* 38: W7–W13. <https://doi.org/10.1093/nar/gkq291>
- Bankevich A, Nurk S, Antipov D, Smoot M, Shumway M, Antonescu C, Salzberg SL (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19: 455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Bao L, Zhang YH, Gu X, Gao YF, Yu YB (2018) The complete mitochondrial genome of *Eterusia aedea* (Lepidoptera, Zygaenidae) and comparison with other zygaenid moths. *Genomics* 111: 1043–1052. <https://doi.org/10.1016/j.ygeno.2018.06.007>

- Bazinet AL, Cummings MP, Mitter KT, Mitter C (2013) Can RNA-Seq resolve the rapid radiation of advanced moths and butterflies (Hexapoda: Lepidoptera: Apoditrysia)? An exploratory study. PLoS ONE 8: e82615. <https://doi.org/10.1371/journal.pone.0082615>
- Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research 27: 573–580. <https://doi.org/10.1093/nar/27.2.573>
- Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsche G, Pütz J, Middendorf M, Stadler PF (2013) MITOS: improved de novo metazoan mitochondrial genome annotation. Molecular Phylogenetics and Evolution 69: 313–319. <https://doi.org/10.1016/j.ympev.2012.08.023>
- Boore JL (1999) Animal mitochondrial genomes. Nucleic Acids Research 27: 1767–1780. <https://doi.org/10.1093/nar/27.8.1767>
- Byun B-K, Bae Y-S (2013) Systematic review of the genus *Yponomeuta* L. (Lepidoptera: Yponomeutidae) in Korea. Insect Koreana 20: 227–237.
- Cameron SL (2014) Insect mitochondrial genomics: implications for evolution and phylogeny. Annual Review of Entomology 59: 95–117. <https://doi.org/10.1146/annurev-ento-011613-162007>
- Cameron SL, Whiting MF (2008) The complete mitochondrial genome of the tobacco hornworm, *Manduca sexta* (Insecta: Lepidoptera: Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths. Gene 40: 112–123. <https://doi.org/10.1016/j.gene.2007.10.023>
- Cao SS, Yu WW, Sun M, Du YZ (2014) Characterization of the complete mitochondrial genome of *Tryporyza incertulas*, in comparison with seven other Pyraloidea moths. Gene 533: 356–365. <https://doi.org/10.1016/j.gene.2013.07.072>
- Cao YQ, Ma C, Chen JY, Yang DR (2012) The complete mitochondrial genomes of two ghost moths, *Thitarodes renzhiensis* and *Thitarodes yunnanensis*: the ancestral gene arrangement in Lepidoptera. BMC Genomics 13: 376. <https://doi.org/10.1186/1471-2164-13-276>
- Chai H-N, Du Y-Z (2012) The complete mitochondrial genome of the pink stem borer, *Sesamia inferens*, in comparison with four other noctuid moths. International Journal of Molecular Sciences 13: 10236–10256. <https://doi.org/10.3390/ijms130810236>
- Chen S, Li F-H, Lan X-E, You P (2016) The complete mitochondrial genome of *Pycnarmon lactiferalis* (Lepidoptera: Crambidae). Mitochondrial DNA Part B 1: 638–639. <https://doi.org/10.1080/23802359.2016.1214551>
- Chen Z-T, Du Y-Z, (2017) The first two mitochondrial genomes from Taeniopterygidae (Insecta: Plecoptera): structural features and phylogenetic implications. International Journal of Biological Macromolecules 111: 70–76. <https://doi.org/10.1016/j.ijbiomac.2017.12.150>
- Clary DO, Wolstenholme DR (1985) The ribosomal RNA genes of *Drosophila* mitochondrial DNA. Nucleic Acids Research 13: 4029–4045. <https://doi.org/10.1093/nar/13.11.4029>
- Coil D, Jospin G, Darling AE (2015) A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. Bioinformatics 31: 587–589. <https://doi.org/10.1093/bioinformatics/btu661>
- Curole JP, Kocher TD (1999) Mitogenomics: digging deeper with complete mitochondrial genomes. Trends in Ecology & Evolution 14: 394–398. [https://doi.org/10.1016/S0169-5347\(99\)01660-2](https://doi.org/10.1016/S0169-5347(99)01660-2)

- Dai L-S, Zhu B-J, Qian C, Zhang C-F, Li J, Wang L, Wei G-Q, Liu C-L (2016) The complete mitochondrial genome of the diamondback moth, *Plutella xylostella* (Lepidoptera: Plutellidae). Mitochondrial DNA Part A 27: 1512–1513. <https://doi.org/10.3109/19401736.2014.953116>
- Garey JR, Wolstenholme DR (1989) Platyhelminth mitochondrial DNA: evidence for early evolutionary origin of a tRNA^{ser}AGN that contains a dihydrouridine arm replacement loop, and of serine-specifying AGA and AGG codons. Journal of Molecular Evolution 28: 374–387. <https://doi.org/10.1007/BF02603072>
- Heikkilä M, Mutanen M, Wahlberg N, Sihvonen P, Kaila L (2015) Elusive ditrysian phylogeny: an account of combining systematized morphology with molecular data (Lepidoptera). BMC Evolutionary Biology 15: 260. <https://doi.org/10.1186/s12862-015-0520-0>
- Katoh K, Rozewicki J, Yamada KD (2017) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings in Bioinformatics. <https://doi.org/10.1093/bib/bbx108>
- Kawahara AY, Breinholt JW (2014) Phylogenomics provides strong evidence for relationships of butterflies and moths. Proceedings of the Royal Society B: Biological Sciences 281: 20140970. <https://doi.org/10.1098/rspb.2014.0970>
- Kim MJ, Jeong JS, Kim JS, Jeong SY, Kim I (2017) Complete mitochondrial genome of the lappet moth, *Kunugia undans* (Lepidoptera: Lasiocampidae): genomic comparisons among macroheteroceran superfamilies. Genetics and Molecular Biology 40: 717–723. <https://doi.org/10.1590/1678-4685-GMB-2016-0298>
- Kim MJ, Kang AR, Jeong HC, Kim KG, Kim I (2011) Reconstructing intraordinal relationships in Lepidoptera using mitochondrial genome data with the description of two newly sequenced lycaenids, *Spindasis takanonis* and *Protantigius superans* (Lepidoptera: Lycaenidae). Molecular Phylogenetics and Evolution 61: 436–445. <https://doi.org/10.1016/j.ympev.2011.07.013>
- Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL (2004) Versatile and open software for comparing large genomes. Genome Biology 5: R12. <https://doi.org/10.1186/gb-2004-5-2-r12>
- Lanfear R, Calcott B, Ho SY, Guindon S (2012) Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular Biology and Evolution 29: 1695–1701. <https://doi.org/10.1093/molbev/mss020>
- Lavrov DV, Brown WM, Boore JL (2000) A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede *Lithobius forficatus*. Proceedings of the National Academy of Sciences of the United States of America 97: 13738–13742.
- Li J, Zhao Y, Lin R, Zhang Y, Hu K, Li Y, Huang Z, Peng S, Ding J, Geng X, Zhang H, Xu Z (2018) Mitochondrial genome characteristics of *Somena scintillans* (Lepidoptera: Erebididae) and comparison with other Noctuoidea insects. Genomics. <https://doi.org/10.1016/j.ygeno.2018.08.003>
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451–1452. <https://doi.org/10.1093/bioinformatics/btp187>
- Lowe T.M, Eddy SR (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research 25: 955–964. <https://doi.org/10.1093/nar/25.5.0955>

- Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* 1: 18. <https://doi.org/10.1186/2047-217x-1-18>
- Mikkel S, Lindgreen S, Orlando L (2016) AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC research notes* 9: 88. <https://doi.org/10.1186/s13104-016-1900-2>
- Mitter C, Davis DR, Cummings MP (2017) Phylogeny and evolution of Lepidoptera. *Annual Review of Entomology* 62: 265–283. <https://doi.org/10.1146/annurev-ento-031616-035125>
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32: 268–274. <https://doi.org/10.1093/molbev/msu300>
- Ojala D, Montoya J, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290: 470–474. <https://doi.org/10.1038/290470a0>
- Perna NT, Kocher TD (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *Journal of Molecular Evolution* 41: 353–358. <https://doi.org/10.1007/BF01215182>
- Rambaut A, Suchard MA, Xie D, Drummond AJ (2014) Tracer v1.6. <http://tree.bio.ed.ac.uk/software/tracer/>
- Regier JC, Mitter C, Zwick A, Bazinet AL, Cummings MP, Kawahara AY, Sohn J-C, Zwickl DJ, Cho S, Davis DR, Baixeras J, Brown J, Parr C, Weller S, Lees DC, Mitter KT (2013) A large-scale, higher-level, molecular phylogenetic study of the insect order Lepidoptera (moths and butterflies). *PLoS ONE* 8: e58568. <https://doi.org/10.1371/journal.pone.0058568>
- Ronquist F, Huelsenbeck JP (2003) MrBayes3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574. <https://doi.org/10.1093/bioinformatics/btg180>
- Sohn JC, Regier JC, Mitter C, Davis D, Landry J-F, Zwick A, Cummings MP (2013) A molecular phylogeny for Yponomeutoidea (Insecta, Lepidoptera, Ditrysia) and its implications for classification, biogeography and the evolution of host plant use. *PLoS ONE* 8: e55066. <https://doi.org/10.1371/journal.pone.0055066>
- Salvato P, Simonato M, Battisti A, Negrisola E (2008) The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera, Notodontidae). *BMC Genomics* 9: 331. <https://doi.org/10.1186/1471-2164-9-331>
- Silvestro D, Michalak I (2012) raxmlGUI: a graphical front-end for RAxML. *Organisms Diversity & Evolution* 12: 335–337. <https://doi.org/10.1007/s13127-011-0056-0>
- Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690. <https://doi.org/10.1093/bioinformatics/btl446>
- Taanman JW (1999) The mitochondrial genome: structure, transcription, translation and replication. *Biochim Biophys Acta* 1410: 103–123. [https://doi.org/10.1016/S0005-2728\(98\)00161-3](https://doi.org/10.1016/S0005-2728(98)00161-3)

- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* 30: 2725–2729. <https://doi.org/10.1093/molbev/mst197>
- Timmermans MJTN, Lees DC, Simonsen TJ (2014) Towards a mitogenomic phylogeny of Lepidoptera. *Molecular Phylogenetics and Evolution* 79: 169–178. <https://doi.org/10.1016/j.ympev.2014.05.031>
- van Asch B, Blibech I, Pereira-Castro I, Rei FT, da Costa LT (2016) The mitochondrial genome of *Prays oleae* (Insecta: Lepidoptera: Praydidae). *Mitochondrial DNA Part A, DNA Mapping, Sequencing, and Analysis* 27: 2108–2109. <https://doi.org/10.3109/19401736.2014.982579>
- van Nieukerken EJ, Kaila L, Kitching IJ, Kristensen NP, Lees DC, Minet J, Mitter C, Mutanen M, Regier JC, Simonsen TJ, Wahlberg N, Yen SH, Zahiri R, Adamski D, Baixeras J, Bartsch D, Bengtsson BÅ, Brown JW, Bucheli SR, Davis DR, de Prins J, de Prins W, Epstein ME, Gentili-Poole P, Gielis C, Hättenschwiler P, Hausmann A, Holloway JD, Kallies A, Karsholt O, Kawahara AY, Koster S, Kozlov M, Lafontaine J D, Lamas G, Landry J-F, Lee S, Nuss M, Park K-T, Penz C, Rota J, Schintlmeister A, Schmidt BC, Sohn J-C, Solis MA, Tarmann GM, Warren AD, Weller S, Yakovlev RV, Zolotuhin VV, Zwick, A (2011) Order Lepidoptera Linnaeus, 1758. *Zootaxa* 3148: 212–221. <https://doi.org/10.11646/zootaxa.3148.1.41>
- Vila M, Björklund M (2004) The utility of the neglected mitochondrial control region for evolutionary studies in Lepidoptera (Insecta). *Journal of Molecular Evolution* 58: 280–290. <https://doi.org/10.1007/s00239-003-2550-2>
- Wei S-J, Shi B-C, Gong Y-J, Li Q, Chen X-X (2013) Characterization of the mitochondrial genome of the diamondback moth *Plutella xylostella* (Lepidoptera: Plutellidae) and phylogenetic analysis of advanced moths and butterflies. *DNA and Cell Biology* 32: 173–187. <https://doi.org/10.1089/dna.2012.1942>
- Wei SJ, Shi M, Chen XX, Sharkey M.J, van Achterberg C, Ye GY, He JH (2010) New views on strand asymmetry in insect mitochondrial genomes. *PLoS ONE* 5: e12708. <https://doi.org/10.1371/journal.pone.0012708>
- Wu Y-P, Zhao J-L, Su T-J, Li J, Yu F, Chesters D, Fan R-J, Chen M-C, Wu C-S, Zhu C-D (2012) The complete mitochondrial genome of *Leucoptera malifoliella* Costa (Lepidoptera: Lyonetiidae). *DNA and Cell Biology* 31: 1508–1522. <https://doi.org/10.1089/dna.2012.1642>
- Wu YP, Zhao J-L, Su T-J, Luo A-R, Zhu C-D (2016) The complete mitochondrial genome of *Choristoneura longicellana* (Lepidoptera: Tortricidae) and phylogenetic analysis of Lepidoptera. *Gene* 591: 161–176. <https://doi.org/10.1016/j.gene.2016.07.003>
- Yang X, Cameron SL, Lees DC, Xue D, Han H (2015) A mitochondrial genome phylogeny of owlet moths (Lepidoptera: Noctuoidea), and examination of the utility of mitochondrial genomes for lepidopteran phylogenetics. *Molecular Phylogenetics and Evolution* 85: 230–237. <https://doi.org/10.1016/j.ympev.2015.02.005>
- Zhang D-X, Hewitt GM (1997) Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary study. *Biochemical Systematics and Ecology* 25: 99–120. [https://doi.org/10.1016/s0305-1978\(96\)00042-7](https://doi.org/10.1016/s0305-1978(96)00042-7)

Supplementary material 1

Supplementary Tables S1–S8

Authors: Mingsheng Yang, Bingyi Hu, Lin Zhou, Xiaomeng Liu, Yuxia Shi, Lu Song, Yunshan Wei, Jinfeng Cao

Data type: molecular data

Explanation note: **Table S1.** List of Lepidoptera species used in phylogenetic analyses.

Table S2. The best scheme and substitution models for the PCG123R dataset.

Table S3. The best scheme and substitution models for for the PCG123 dataset. **Table S4.** AT-skew and GC-skew of the *Yponomeuta montanatus* mitogenome.

Table S5. A + T content (%) in three codon positions in mitochondrial protein-coding genes of reported yponomeutoid mitogenomes. **Table S6.** Codon usage in mitochondrial protein-coding genes of the *Yponomeuta montanatus*. **Table S7.** Start and stop codons of mitochondrial protein-coding genes of four yponomeutoid species. **Table S8.** Evolutionary rates of mitochondrial protein-coding genes among reported yponomeutoid mitogenomes.

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: <https://doi.org/10.3897/zookeys.879.35101.suppl1>

Supplementary material 2

Supplementary Figures S1–S3

Authors: Mingsheng Yang, Bingyi Hu, Lin Zhou, Xiaomeng Liu, Yuxia Shi, Lu Song, Yunshan Wei, Jinfeng Cao

Data type: molecular data

Explanation note: **Figure S1.** ML trees inferred from RAxML method based on PCG123 (A) and PCGAA (B) datasets. Number on node represents bootstrap replicate. **Figure S2.** ML trees inferred from IQ-TREE method based on PCG123 (A) and PCGAA (B) datasets. Number on node represents bootstrap replicate. **Figure S3.** BI trees inferred from MrBayes method based on PCG123 (A) and PCGAA (B) datasets. Number on node represents posterior probability.

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: <https://doi.org/10.3897/zookeys.879.35101.suppl2>